人工智能时代信息价值开发的伦理约束机制研究

杨旦修
（云南财经大学传媒学院，云南 昆明 650221）

摘　要：人工智能持续发展的同时其技术所诱发的数据伦理和算法伦理风险成为当下热议的话题。数据伦理风险问题主要包括个人隐私安全、信息安全、数据权利边界模糊和智能鸿沟等问题；算法伦理风险问题主要包括算法偏见、算法歧视以及算法推荐隐含的负面诱导和信息茧房风险。这些伦理风险亟待成熟的约束机制进行解决，主要可以从以数据权利的界定为核心的法律法规的出台；主体责任、行业自律以及公民网络素养相结合的伦理机制的构建以及算法程序的技术创新三个维度进行探讨。

关键词：人工智能；数据伦理；算法伦理；约束机制　　**中图分类号：**TP392　　**文献标识码：**A

文章编号：1671-0134（2021）07-068-05　　**DOI：**10.19483/j.cnki.11-4653/n.2021.07.019

本文著录格式：杨旦修 . 人工智能时代信息价值开发的伦理约束机制研究 [J]. 中国传媒科技，2021（07）：68-72.

1. 信息价值开发的人工智能伦理风险的生成逻辑

“1956 年，在美国达特茅斯召开达特茅斯专题研讨会上，以明斯基、麦卡锡、纽厄尔为首的一批年轻科学家们共聚一堂，一起研究和探讨有关机器模拟智能技术的话题。”[1] 此次会议标志着人工智能这门新兴科学技术的诞生。

人工智能的数据和算法分析为人类的生产生活提供技术支撑的同时也诱发了诸多的数据伦理和算法伦理问题，可能会影响到社会的稳定和公平。例如一些犯罪分子可能会非法窃取他人的隐私数据开展犯罪活动，如暗网上的非法人口交易；不加约束的算法系统可能会对特定的群体产生算法偏见等数据伦理和算法伦理风险。

1.1　数据伦理风险的生成逻辑

简单来说，数据伦理风险涉及由大型的数据收集和开发引起的伦理问题，数据伦理风险主要包括个人隐私在内的信息安全问题、数据霸权下的数据权利的受损以及信息垄断等造成的智能鸿沟问题。

1.1.1　生成因素

一是数据来源的复杂性。数据是对网络用户浏览和使用痕迹的记录，一定程度上带有数据消费者和收集者的主观偏见；算法程序中训练数据量的多少同样也会影响数据输出的价值导向。例如 FBI 在 2016 年发布的犯罪统计数据中，黑种人犯罪率达到总犯罪量 408871 起的 37%， 明显已经高于白人，但是究其原因，长期歧视带来的暴力倾向本就尚存争议，更何况群体的特质不能以偏概全代表整个黑人群体。

二是数据权利边界的模糊。平台积聚的用户数据具体主要有用户主动留存的浏览痕迹的数据，还包括用户在进行信息消费时附加的一些条件而产生的被动型数据，这些附加条件一定程度上是为了规避用户的知情同意权，是平台和应用主体之间缺乏伦理规范的界定和约束的表现。例如 2020 年 2 月 17 日，新华社披露了春节期间不法分子通过大数据手段，泄露和转载武汉返乡人员的大量个人信息，造成了多人数据信息安全受到威胁。

三是数据应用市场的失衡。人工智能的大数据开发和应用就国内市场而言，大数据开发和应用仍然处于比较混乱的阶段，数据的透明化、公开化发展方面还存在着非良性的竞争形势，大数据发掘的隐蔽性和去隐私化之间，还没有合理有效的市场制约机制的出现。

比如 2018 年 7 月，大众点评通过爬虫技术抓取了大量小红书用户的个人信息，对小红书平台上用户的原创笔记进行抄袭和搬运，互联网企业非法盗取用户数据不仅会对互联网用户的个人权利造成侵害，而且会导致数据应用市场失衡，失去良性循环的发展空间。

1.1.2　生成机制

数据来自个体，个体深植于文化内涵之中，文化意识本身带有主观性，一旦整个社会对某个少数群体存在结构性的偏见，这些偏差也必然地会反映在数据上；这是就数据来源而言，而在平台对这些数据进行开发和应用的过程中，数据群里边界的界定又是一个问题，用户让渡部分权利获取平台提供的信息服务，而部分平台缺乏数据伦理规范的数据滥用行为越来越多，数据市场的自发性和趋利性洞察到这一先机，更多的平台通过不正当竞争的手段扒取用户数据，造成了数据应用市场的失衡（图 1）。

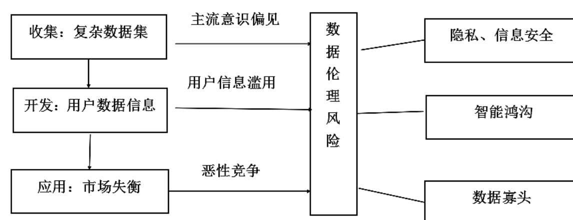


图 1　数据伦理风险生成机制示意图

1.1.3 数据伦理风险分类

首先是隐私安全问题。在数据伦理风险的探讨中，隐私问题是数据伦理问题的核心，人工智能技术的发展以大数据的收集为依托，而大数据的收集过程中也会发生对用户个人信息二次利用、非法收集和信息泄露和扩散等问题。2018年9月新华社报道称以9521开头的电话诈骗团伙一般会通过各种互联网渠道非法收集个人详细信息，用户个人受到的骚扰诈骗电话号码数辐射人群达410万之多，给用户的信息安全造成了威胁。

其次是信息安全问题。随着互联网的普及和人工智能技术的发展，个人的信息存在着随时被泄露的风险。2020年7月北京互联网法院做出一审判决，认定微信读书和抖音两款App均存在对用户个人信息的侵害情形，大数据在整合抓取个人信息时很容易泄露个人的相关信息，人们仿佛置身于楚门的世界，在网站和社交媒体上的一言一行都存在被监控的风险。

再次是用户数据权利受损问题。“根据数据权利的内涵，数据的收集和使用需要征得用户的知情同意，无序无度的数据共享只会导致数据主体权利的受损和数据征集秩序的混乱。”[2] 数据权利包括知情同意权、隐私权、删除权和数据携带权等人的基本权利在数据领域的具体化。2016年的4月份，黑客对土耳其国家信息安全库进行攻击，造成包括土耳其总统埃尔多安在内的5000万名公民的信息在网络上泄露，而信息安全部门并未对个人的敏感数据和个人信息访问权限进行合理保护。

最后是数据垄断和数字鸿沟问题。针对数据的开发者而言，开发者对数据的垄断会导致权益分配不均和数据侵权行为。国外的Facebook、Google，国内的腾讯、阿里巴巴等商业巨头掌握着广泛的信息源和技术优势，这些巨头们垄断了数据应用市场的绝大部分数据资源。对数据的消费者来说，相比没有机会接触网络的人群，对更频繁地接触网络的用户，能够接收到网络运营商提供的有效信息也就越多。

根据《2019年全球互联网数字报告》显示，截至2018年12月，全球的互联网用户已经达到了41亿，尚有48%左右的人群未被推广和普及互联网络，这也就意味着面对庞杂的数据市场，人们并非享有公平的数据分配权利，甚至使用网络和不使用网络的人群之间的知识沟将会越来越大。

1.2 算法伦理风险的生成逻辑

“算法伦理则风险是通过特定的指令和程序对数据进行开发时，产生的不可预见或者潜在的不良后果。”[3] 主要原因有算法的不可控性、算法开发和应用过程中的主观价值偏好和伦理约束机制尚未完善，存在的风险主要有算法推荐可能引发的负面诱导 、信息茧房问题、算法偏见以及算法歧视问题。

1.2.1 生成因素

一是算法的不可控性。机器学习可以通过合适的算法和足够的训练数据自主的编写代码，利用数据产生新的模式和知识，基于大数据系统可以对决策的规则进行有效的修改、延伸和调整。不同数据之间的规模、质量和来源各不相同，它们也在影响着算法程序的复杂性，因此算法所衍生出的部分风险存在着不可预测性。

2016年，Facebook平台就因通过算法将1972年拍摄的《战火中的女孩》反战作品确定为不良信息而被查封和禁用，这让Facebook也受到了巨大舆论压力，平台在算法程序的自主调节和引发的伦理风险之间难以平衡，陷入挑战。

二是算法开发和应用过程中的主观价值偏好。算法决策在更多时候是用以往的数据预测未来的趋势走向，算法的模型设计和数据的输入都影响着预测的结果，开发设计者可能会将自己的主观意志嵌入到算法系统。对用户而言，算法开发成功后，用户在使用智能设备时可以根据自己的客观需求自主设置参数。

《算法之美》一书中提到一个由于个人主观价值偏好导致数据分析有误的过度拟合的例子，他们通过详细的参数指标建立了一个调查问卷，由于选用详细的个人意见相关的参数指标导致错过了采样数据点而起不到预测的效果，算法程序的设计亦是如此，过度受到设计者和用户主观价值偏好的影响也会引起不必要的风险。

三是尚未完善的伦理约束机制。算法可以自主调节程序和参数，这也给算法程序增加了不确定性，另一方面，人们对于算法程序直接的责任主体和可控机制尚且缺乏进一步的探讨和研究。作为算法开发和应用的平台主体，如果缺乏完善的伦理机制制约，将会影响整个互联网生态环境的良性循环，诱发更多的算法伦理风险。例如2018年4月，字节跳动公司因旗下产品“内涵段子”内容导向低俗和价值导向问题受处罚并永久性关停，其子产品“抖音”平台的用户评论也被清除。

1.2.2 生成机制

在算法的开发和设计过程中，机器学习技术等人工智能技术通过合适的算法和训练数据可以自主编写代码，执行常见的信息服务的指令，但是由于缺乏对人类社会文化内涵的深层分析和价值认同，很难面面俱到，常常会超出算法和平台设计者的伦理规制，衍生伦理风险；应用过程中，设计者的主观意志以及个人用户对服务参数的自主修改等因素都会进一步对收集的数据和新的算法程序的规范产生误差影响，此外平台算法机制引发的

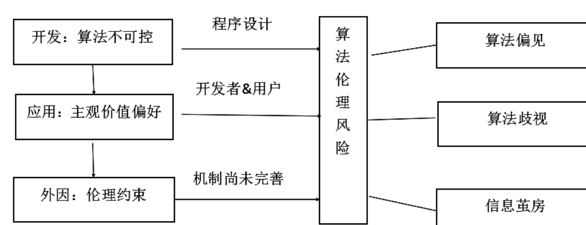


图2 算法伦理风险生成机制示意图

价值导向和内容导向都缺乏伦理法规全面的约束。

1.2.3 算法伦理风险分类

首先是算法偏见问题。“算法偏见是指算法在进行决策时建立起来的对人、事、物的消极认知和态度。”[4]《Nature》杂志将这种在“黑匣子”面前无法了解决策过程，只能明白结果的形式称为“偏见进，偏见出。” 比如 2015 年 Google 的图像识别软件之前将黑人“猿猴”或者“黑猩猩”等动物进行错误地标记引发了一系列的种族歧视争议。

其次是算法歧视问题。算法歧视在日常生活中最常遇到的就是大数据杀熟的问题，大数据杀熟主要是指经营者可以在售卖同一项产品或者服务的过程中，对待不同的消费者采取不同的售卖标准，从而赚取更多的消费者剩余的行为。比如 2020 年 12 月份，天猫超市就被多名消费者举报天猫 88 会员用户的商品的价格显示要比一般用户看到的商品价格高出很多。

最后是算法推荐可能引发的负面诱导和信息茧房问题。基于大数据系统的精准算法推荐可能存在着负面诱导的风险。比如 2018 年 3 月 20 英国大数据公司剑桥分析在总统大选期间与美国总统特朗普进行合作，非法获得 Facebook 平台 5000 万个用户的资料，通过详细的数据信息对用户实施带有指向性的选举信息推送，干涉全球一些国家的选举，存在负面诱导用户选举行为的嫌疑。

信息茧房则是随着智能化传播而产生的一种现象。对一般传播范畴，它主要是指平台的技术逻辑导致对用户信息推送的单一和片面，长此以往逐渐导致用户的自我封闭现象和信息窄化行为的发生。算法的精准推荐确实实现了基于用户画像的用户准需求对接，但是它大大加速了用户与平台之间、用户与用户之间的信息不对称，容易导致单一信息对消费者的精神控制。

2. 信息价值开发的数据伦理约束

数据伦理问题包括数据的挖掘存储以及数据的利用两个方面，由于数据来源的复杂性、数据应用市场的失衡以及数据权利边界的模糊，极易引发信息安全、数据霸权、信息鸿沟在内的数据伦理问题，因此有必要对其进行数据来源的约束、数据权力边界的约束和数据应用市场的约束。

2.1 数据来源的约束

数据具有不确定、动态性和不完全的特点，但是在算法程序设计中的训练数据较为简单，且是以静态的形式分布与采样信息中，因此采集的数据就存在一定的局限。另一方面，数据是对网络用户浏览和使用痕迹的记录，一定程度上带有数据消费者和收集者的主观偏见。例如，机器学习算法没有考虑攻击者会有意识地喂给机器学习系统一些恶意数据，从而导致训练得到的模型偏离了真实情况。由于人工智能涉及的相关技术很多，因此，这类的安全问题将是非常广泛存在的。

2.2 数据权利边界的约束

数据权力边界的界定会根据具体的场景而发生变化，在变化的过程中又会有新的权利界定问题的产生，上述的问题加剧了数据权利的归属问题的难度。大部分平台积聚的用户数据主要可以分为两类：一类是用户主动留存的数据，比如通过智能搜索引擎进行的信息查询和订阅等。另一类是用户在进行信息消费时附加的一些条件而产生的被动型数据，比如用户登录手机 App 时被平台要求获取位置、通讯录等信息的访问权限，这些附加条件一定程度上是为了规避用户的知情同意权，降低数据获取的门槛而设置的，是平台和应用主体之间缺乏伦理规范的界定和约束的表现。

2.3 数据应用市场的约束

“人工智能的大数据开发和应用就国内市场而言，大数据开发和应用仍然处于比较混乱的阶段，数据的透明化、公开化发展方面还存在着非良性的竞争形势，大数据发掘的隐蔽性和去隐私化之间，还没有合理有效的市场制约机制的出现。”[5] 如视觉识别领域的人脸识别技术被广泛应用推广，其收集的用户数据未得到妥善处理，可能就会损害用户的利益。另一方面，用户在互联网上产生的浏览痕迹又会成为新的数据信息进行沉淀和又一次收集，在这个过程中也会产生用户个人隐私数据泄露的风险。

3. 信息价值开发的算法伦理约束

算法伦理涉及人工智能、机器学习和机器人如何解释算法，它关乎人们面临道德困境时是否能够为其提供正确的决策，对算法的约束主要包括不可控性、算法开发和应用过程中的主观价值偏好以尚未完善的伦理约束机制的约束。

3.1 算法的不可控性的约束

“机器学习可以通过合适的算法和足够的训练数据自主编写代码，利用数据产生新的模式和知识，基于大数据系统可以对决策的规则进行有效的修改、延伸和调整。”[6] 不同数据之间的规模、质量和来源各不相同，它们也在影响着算法程序的复杂性，因此算法所衍生出的部分风险存在着不可预测性。

3.2 算法开发和应用过程中的主观价值偏好的约束

“算法决策在更多时候是用以往的数据预测未来的趋势走向，算法的模型设计和数据的输入都影响着预测的结果，算法本质上是数学式的或者代码式的意见表达手段。”[7] 对设计者而言，算法设计、数据筛选和应用标准等因素都隐含着算法开发者的主观态度，开发设计者可能会将自己的主观意志嵌入到算法系统；对用户而言，算法开发成功后，用户在使用智能设备时可以根据自己的客观需求自主设置参数。

3.3 伦理机制的构建和约束

算法可以自主调节程序和参数给算法程序增加了不确定性，另一方面，人们对算法程序直接的责任主体和

可控机制尚且缺乏进一步的探讨和研究。

霍金 、施密特等科学家之前都警惕强人工智能或者超人工智能可能威胁人类生存。从阿西莫夫提出的机器人三定律到 2017 年阿西洛马会议提出的 23 条人工智能原则，人们针对人工智能可能存在的伦理安全问题的谈论从不间断，但是对于能够结合当前算法程序发展的实际情况，对主体责任机制进行有针对性的探讨的理论机制还并不多。因此面对数据伦理和算法伦理带来的新的风险与挑战，建立一套科学完整的伦理约束机制显得尤为必要。

4. 信息价值开发的人工智能伦理约束的价值

信息价值开发的人工智能伦理约束的价值主要可以分为三个方面，保护个人数据安全，尊重人的权利；更好地发挥大数据技术和算法程序的积极作用，把握社会信息安全的边界；引导信息价值伦理观念的重塑，促进人的全面发展。

4.1 保护个人数据安全，尊重人的权利

一切都将被记录，一切都将被分析，消极的伦理观念放大了数据和算法对个人隐私安全的冲击，算法的目标是人类设计的，算法本身并不具备设定毁灭人类的主观意志。我们应该关注的是如何是数据和算法程序开发在合乎伦理规范的设计者手中更上一层楼，为人类谋福祉。

“从个人层面来看，人工智能时代的信息安全包含着自由、尊严和伦理价值。”[8] 智能时代用户的个人信息数据和自由和权利密切相关，由于自由、尊严等个人权利是公认的伦理价值，而个人隐私能否得到保护又蕴含的自由和尊严等个人权利是否得到尊重，因此，对于智能时代个人的信息安全的保护，在个人层面上具有重要的伦理价值。

4.2 更好地发挥大数据技术和算法程序的积极作用，把握社会信息安全的边界

实际上这些伦理风险与人工智能带来的技术红利是相伴而生的，从社会层面看，人工智能的信息安全对实现文明、和谐等伦理价值有着积极意义。文明与和谐是社会存在和发展的应有之义，是获得道德的正当性辩护的有效价值目标。比如，拥有海量数据的现代社会组织，极易成为黑客的攻击对象，数据泄露的风险，会对企业的声誉和社会稳定产生不良影响。那么信息安全的边界将会界定得更为明晰。相反情况下，信息安全的阙如有伤于社会的文明、和谐。

4.3 引导信息价值伦理观念的重塑，促进人的全面发展

“小数据时代的伦理观念对于当今人工智能开发的伦理风险的制约不具有普适性，这也启示我们要树立公平正义、开放多元、互助共享的新型伦理观念。”[9] 一方面是数据与算法的开发者要重视对正在完善的相关法律法规和行业规范的遵守，另一方面数据的主体应该提升自己的数据能力和网络素养，通过以人为本的数据权利追求新的信息自由，实现互利共赢。通过有效的数据伦理和算法伦理约束机制，引导用户对多元信息都能有选择的接触和理解，防止出现信息偏食而对人的全面发展产生影响。

5. 信息价值开发的智能伦理风险挑战和约束机制

信息价值开发的智能伦理风险挑战，首先是个人隐私保护的界定问题；其次是算法推荐机制的设计问题；最后是数据信息选择的复杂性问题。针对上述问题，首先要加强个人隐私信息的保护；其次要构建算法内外部约束机制；最后是提高人工智能技术创新能力。

5.1 信息价值开发的智能伦理风险挑战

首先是个人隐私保护的界定问题。不同的国家、不同的民族对隐私问题存在不同的定义，联合国教科文组织于 2016 年通过的《关于机器人伦理的初步草案报告》指出，对智能技术及其衍生产品对人类造成的损害可以通过产品责任的相关制度进行调整，在这里，智能产品主要被作为一般的科技产品对待。但是这些意见仅仅只针对弱人工智能的智能产品，对未来能够拥有独立意识的智能科技产品就不具有普适性。

其次是算法推荐机制的设计问题。不同用户对信息的需求指向不同，算法推荐会为一些高强度的生活和工作节奏的用户节约时间和选择成本，但是也会对需要汲取多元信息数据的用户带来困扰。从技术创新方面来看，技术的开发和应用阶段仍然存在较多技术风险，比如个人隐私安全的保护以及数据偏见等问题。

最后是数据信息选择的复杂性问题。由于用户产生的数据信息存在复杂性，有些无用的数据会被保留，而重要的数据会被删除，这就给数据清洗工作带来了挑战。数据开发者的主体责任约束机制有待完善，机构之间的数据开发缺乏自律规范；消费者的权利观念比较淡薄。

5.2 信息价值开发的伦理约束机制

第一是完善数据开发和算法程序规范相关的法律法规和人工智能的监管体系，加强个人隐私安全和信息安全的保护。在 2018 年 5 月，欧盟颁布了《一般数据保护条例》（GDRB），该条例对个人数据的获取权、修改权、被遗忘权、异议权等进行了明确的界定，通过有针对性的法律法规的制定可以有效的对数据开采和算法程序进行规范，主动约束算法使用者自身的行为，这也是建构算法伦理的重要进路。

完善法律法规的同时，应该建立相关的监管标准和行业规范，在实践中合理的配置道德、教育、艺术、宗教等不同内容的权重，使得人工智能设计系统可以与人类社会的道德标准保持协调统一。

第二是构建算法治理的内外部约束机制，增强算法程序的价值敏感设计。计算机系统的算法偏见主要可以分为预存偏见、技术偏见和突生性偏见。“预存偏见源

自社会制度、态度和习惯。”[10] 比如我们在百度的搜索栏输入“首席执行官”等关键词，那么界面显示的人物一般是男性；技术性偏见主要是由于技术限制了程序的开发与设计；而突生性主要是社会文化、价值理念的冲突等引起的，比如微软在推特的人工智能聊天机器人 Tay 在与用户聊天时，受到部分用户带有种族歧视色彩的言论的影响造成习得偏见这样的现象。

就算法程序的设计者和开发者而言，“需要加强算法设计者道德责任的建设，提高道德想象力的构建能力。”[11] 在算法的开发阶段能够通过对各种既定情景的设想可能带来的后果进行风险评估，最后通过这些评估经验做出有效的决策标准，将道德观念嵌入整个算法体系的开发和应用的全过程中，定量和定性分析的角度找到价值算法价和值冲突点的平衡点。

第三是提高人工智能技术创新能力，依据科学原理对数据清洗问题进行治理。“开发者要坚持基本的伦理规范，将安全性、有益性、包容多元性等原则贯穿设计始终，明确责任。”[12] 不将人工智能系统用于非法或违反伦理的目的，明确人工智能设计、研发、运营、制造和服务等各个环节中主体的权利和义务，明确伦理风险发生时道德主体应该承担的责任。

首先在数据采集阶段按照行业标准对输入的数据集应该做好数据筛查、数据清洗以及数据格式化的工作，避免数据偏见带来的算法决策的不公平问题。

其次在数据的开发阶段应该摒弃过度攫取数据的数据霸权观念，依照市场机制和行业标准采之有度。“数据开发平台要切实履行主体责任，通过加强技术引领，推动网络巨头自律。”[13] 互联网企业要加强广告主及其用户数据的储备、管理与传输的规范建设与技术保护，坚守数据使用与广告传播的职业伦理、职业道德与职业精神，既能维护企业自身合法权益，又能维护用户权益，责无旁贷地夯实维护国家网络主权、安全与发展利益的基础。

最后的应用阶段，“应该提高算法设计阶段的透明性，以技术支持降低个人信息泄露的风险，倡导公平公开和包容的算法机制，构建双向互动的反馈机制。”[14] 以便平台实时收集用户的信息诉求，形成双向互利的、可持续性的良性循环。媒

参考文献

[1] 黄河燕 . 新工科背景下人工智能专业人才培养的认识与思考 [J]. 中国大学教学，2019（2）：20-25.

[2] 邓国民，李梅 . 教育人工智能伦理问题与伦理原则探讨 [J]. 电化教育研究，2020（6）：39-45.

[3] 宋吉鑫，魏玉东，王永峰 . 大数据伦理问题与治理研究述评 [J]. 理论界，2017（1）：48-54.

[4] 张贵红 . 价值敏感设计与大数据伦理 [J]. 伦理学研究，2019（2）：114-119.

[5] 李伦 .“楚门效应”：数据巨机器的“意识形态”——数据主义与基于权利的数据伦理 [J]. 探索与争鸣，2018（5）：29-31.

[6] 贾红琳 . 大数据算法伦理研究 [D]. 长沙：湖南师范大学，2019.

[7] 李伦 .“楚门效应”：数据巨机器的“意识形态”——数据主义与基于权利的数据伦理 [J]. 探索与争鸣，2018（5）：29-31.

[8] 吕耀怀 . 大数据时代信息安全的伦理考量 [J]. 道德与文明，2019（4）：84-92.

[9] 袁雪 . 大数据技术的伦理“七宗罪”. 科技传播 [J].2016（4）：89-90.

[10] 成素梅，张帆 . 人工智能的哲学问题 [M]. 上海：上海人民大学出版社，2020：185—186.

[11] 段伟文 . 面向人工智能时代的伦理策略 [J]. 当代美国评论，2019（1）：24-38+120.

[12] 李伦，孙保学，李波 . 大数据信息价值开发的伦理约束：机制框架与中国聚焦 [J]. 湖南师范大学社会科学学报，2018（1）： 1-8.

[13] 李伦， 李波 . 大数据时代信息价值开发的伦理问题 [J]. 伦理学研究，2017（05）：100-104.

[14] 闫坤如 . 人工智能的道德风险及其规避路径 [J]. 上海师范大学学报（哲学社会科学版），2018（2）：40-47.

作者简介：杨旦修（1972-），男，湖南邵阳，云南财经大学传媒学院副教授，博士，硕士研究生导师，研究方向：智能传播。

（责任编辑：张晓婧）